

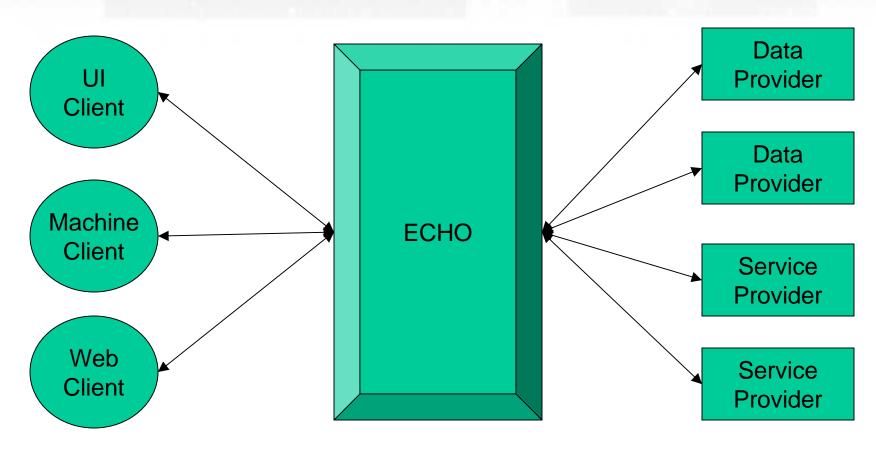
ECHO Operations Workshop

Keith Wichmann ECHO Development Project Manager

June 6-7, 2002



High Level ECHO Context





ECHO Interfaces

Java Toolkit

Client Connector

. . .

Client Interface (RMI) (ECHO API Services)

SOAP to RMI Conversion

Business Objects

ECHO

Catalog

Ingest (XML via FTP)

Extension Extension

XML to V0/ODL

...

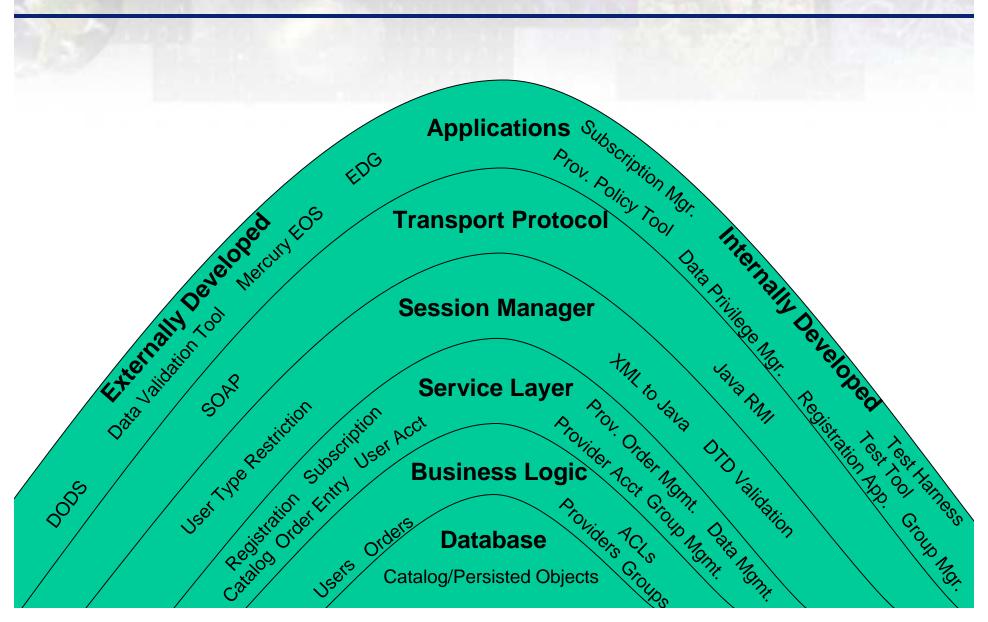
(Quote, Submit, Cancel, Status)

Provider Interface

(SOAP)



ECHO's Layered API Architecture



Overview of ECHO Status

- Version 4.3 has been externally tested
- Version 4.5 is currently being externally tested
 - Looking at 4-6 weeks before its accepted
 - Adds access control lists with group management, performance improvements for catalog service, ...
- Version 5.0 is under development
 - Adds Provider Profile Service, provider controlled delete capabilities, registered user role capability (provider role, admin role), stateless catalog service transactions, SSL, ingest error checking, ...

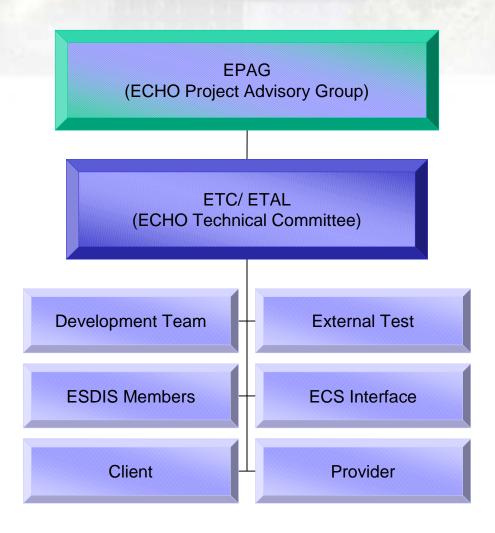


ECHO Operational Timeline

- Current version (4.3) is available for provider ingest and client access
- New hardware should be here
- New version (4.5) available soon
- Current data being operationally tested
 - Operational mode: GSFC (subset), and ORNL
 - Test mode: EDC and NSIDC
- Ops team is coming up to speed
- Need operational acceptance testing
- Client developers need real data!



ECHO Committees





ECHO Preliminary Performance Results

ECHO's original performance design benchmarks:

- Ingest 60,000 granules per day
- 120 queries per hour
- 1000 results per query

Ingest

- 11 minutes for every 1000 granules
- Adequate for 120,000 granules per day

Query

- Roughly 14-17 seconds per query (round trip over the net)
- Adequate for 200 queries per hour

Present

- Assumption: Most queries will only display some small fraction of results (assume 20)
- Roughly 30 seconds per present of 20 granules with all metadata presented
- Adequate for 120 per hour given these assumptions



Ingest Statistics

Split the large file into smaller files

- About 1 second to split a file into 11 smaller files that contains no more then 1000 granules

For about every 1000 granules:

- Clean up processing directory and temporary input staging tables: about 18 seconds
- Decompose: about 42 seconds
- Loading: about 2.5 minutes
- Update tables: about 7.7 minutes
- Clean up temporary input staging tables: about 18 seconds
- Total: about 11 minutes



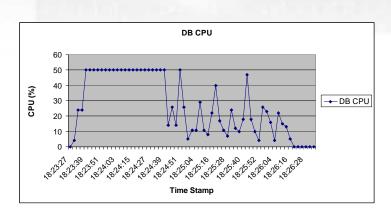
Sample Query Performance

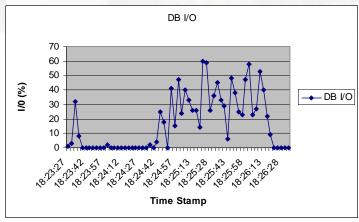
| Queries | Time (sec) | Time/Query |
|---------|------------|-------------|
| 12 | 181 | 15.08333333 |
| 120 | 1742 | 14.51666667 |
| 36 | 604 | 16.7777778 |
| 84 | 1198 | 14.26190476 |

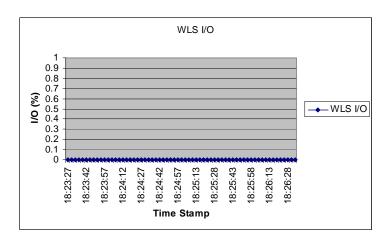


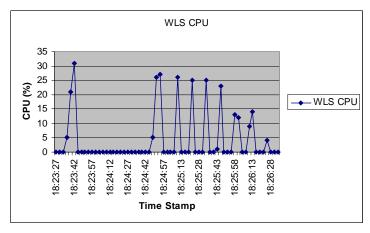
- 12 Queries run sequentially by one user
 - Mixture of query types
- Execution Time: 3:01







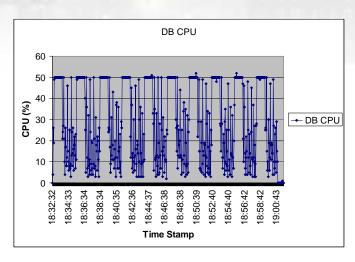


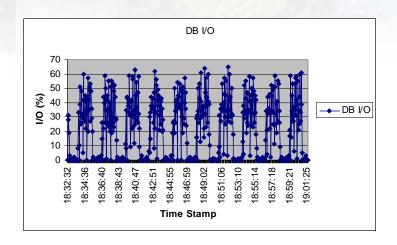


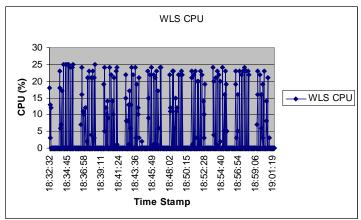


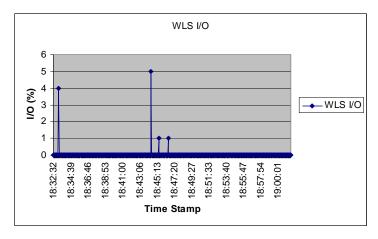
- 12 Queries executed 10 times sequentially, essentially 120 queries
- Execution Time: 29:02







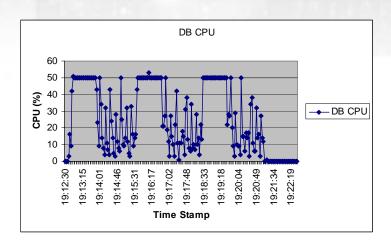


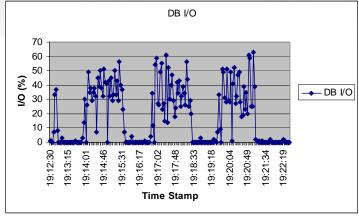


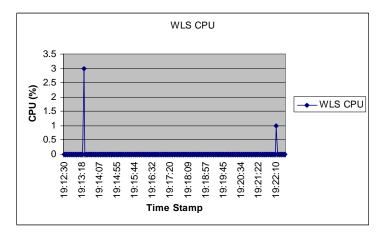


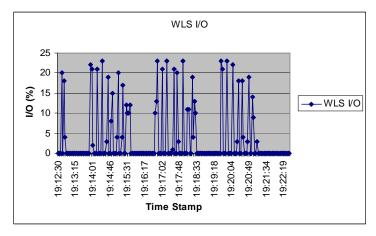
- 36 Queries executed sequentially
- Execution Time: 10:04







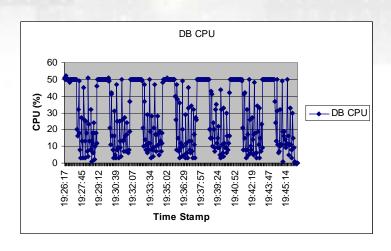


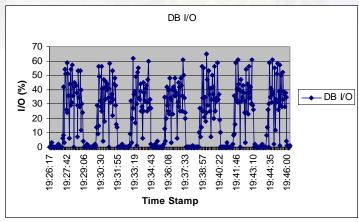


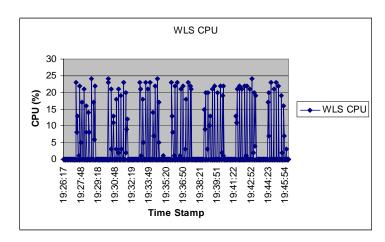


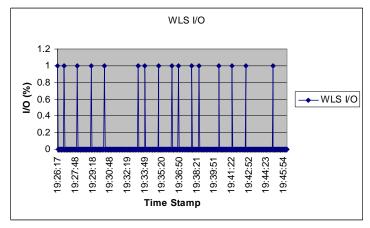
- 84 Queries executed sequentially
- Execution Time: 19:58













Present Performance

- 10 users
- 1 query per user
- Restrictions placed on collection such that some data, but not all would be hidden from user
- Each present statement pulls from a different part of the result set

| Presents | Time (sec) | Time/Present |
|----------|------------|--------------|
| 10 | 300 | 30 |
| 120 | 3600 | 30 |
| 40 | 1200 | 30 |
| 80 | 2400 | 30 |

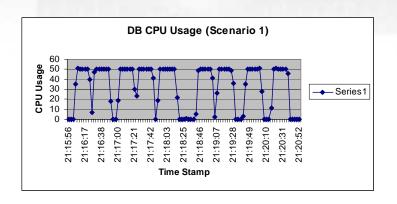


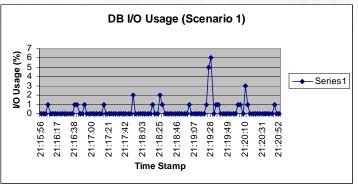
Present - Scenario #1

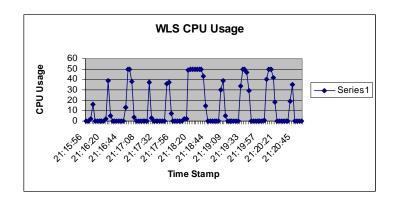
- 10 Present Statements, executed sequentially
- 5 minutes to execute

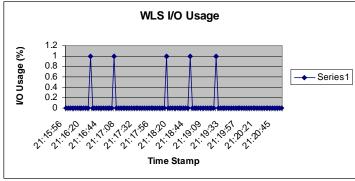


Present - Scenario 1









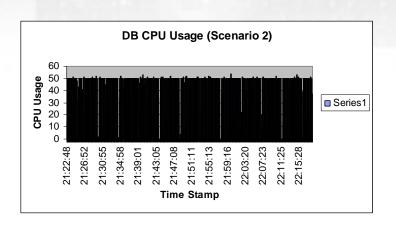


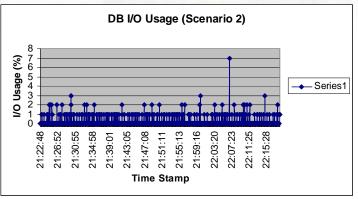
Present - Scenario #2

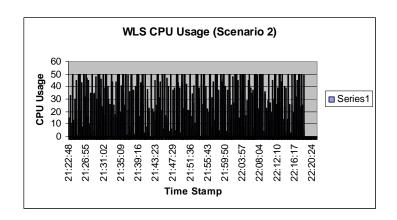
- 120 Present Statements, executed sequentially
 - 10 presents, executed 12 times
- 1 hour to execute

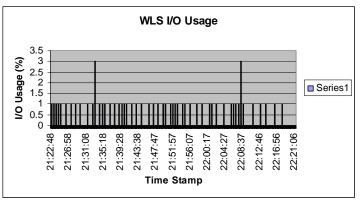


Present - Scenario 2









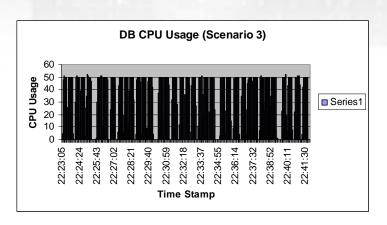


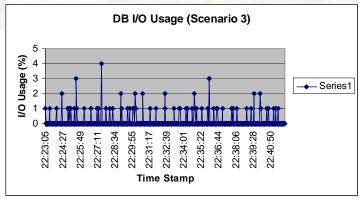
Present - Scenario #3

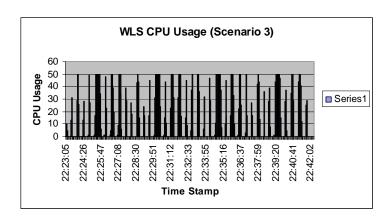
- 40 Present Statements, executed sequentially
- 20 minutes to execute

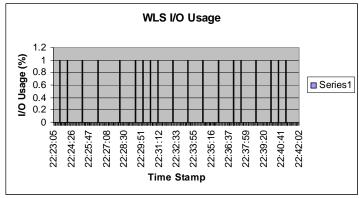


Present - Scenario 3









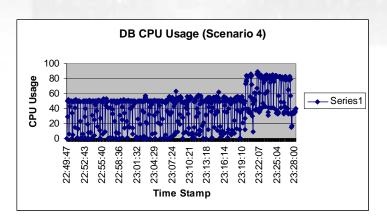


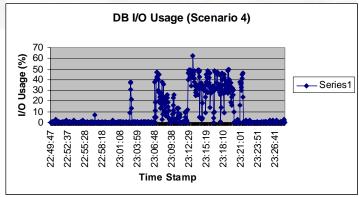
Present - Scenario #4

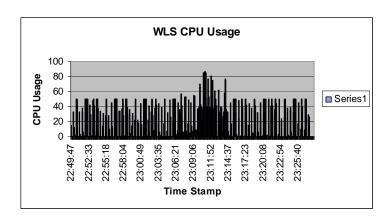
- 80 Present Statements, executed sequentially
- 40 minutes to execute

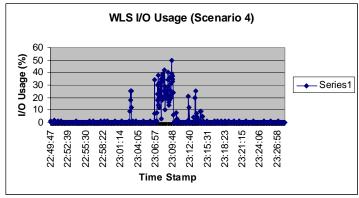


Present - Scenario 4

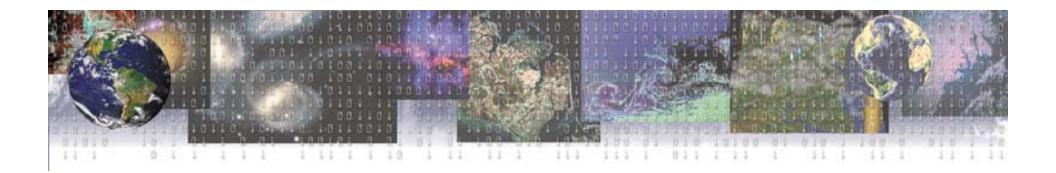










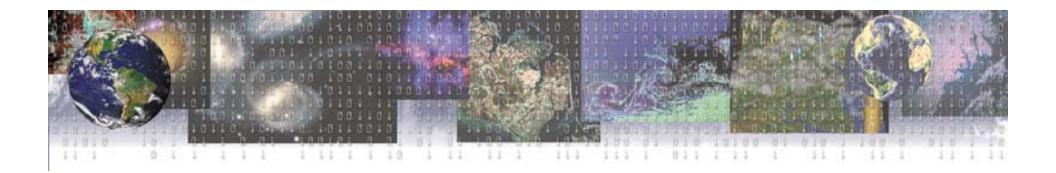


RMA Issues



Hardware (Current)

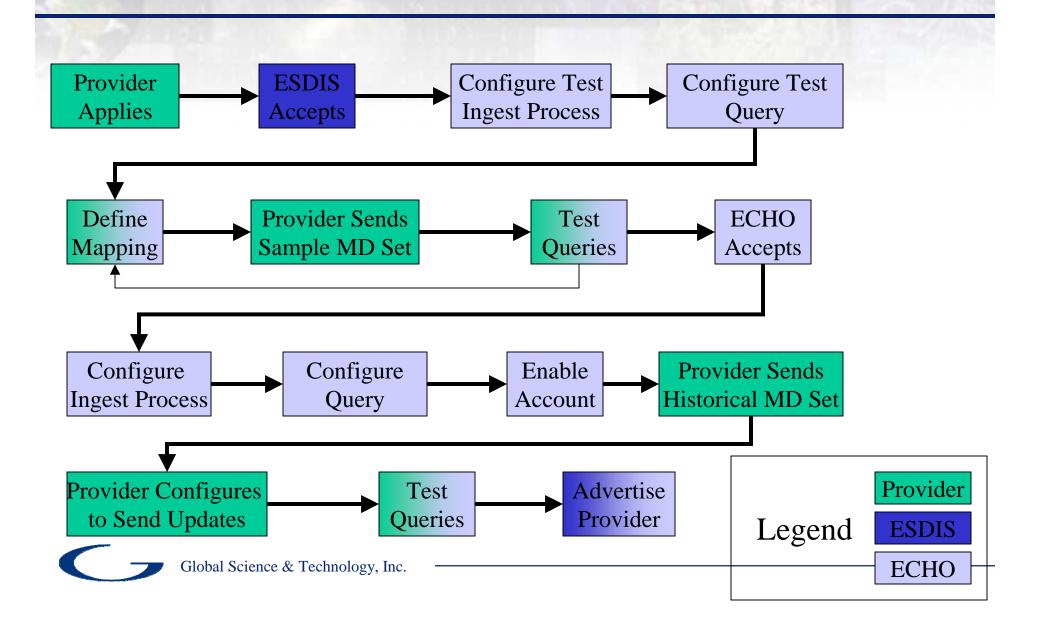
Hardware (Planned)

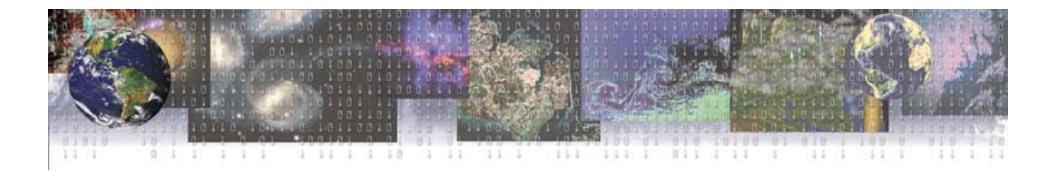


Adding a New Provider



New Provider Process





Provider Services Issues

Policies Access Control User Registration



Iteration 4.7 Provider Policies

Communication Settings

- Configuration information that allows ECHO to effectively communicate certain order transactions from the users to the providers.

Valid for the following transactions:

- Submit, Validate, Quote, Cancel

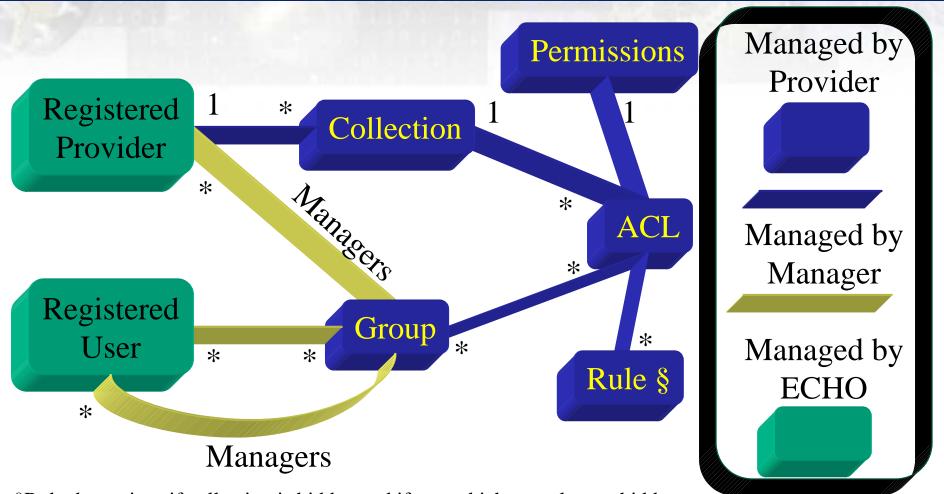
Allowable communication settings for these transactions:

- supported, retryWait(4.4), retryAttempts(4.4), and messageType
- If messageType is SOAP, then also: url, targetobject, and targetURI.
- If messageType is ECS, then also: IP address, and port.

Are there more desired?



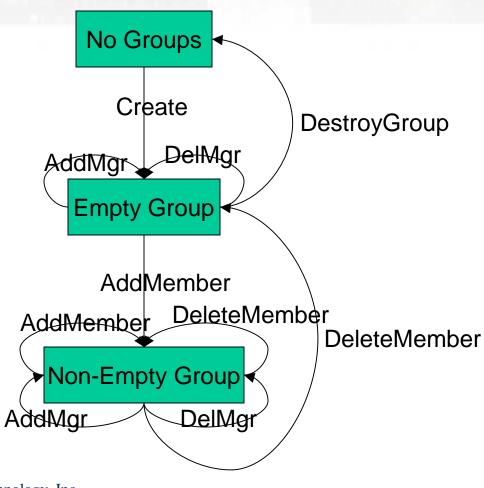
Metadata Visibility - Conceptual Model



§Rule determines if collection is hidden and if not, which granules are hidden

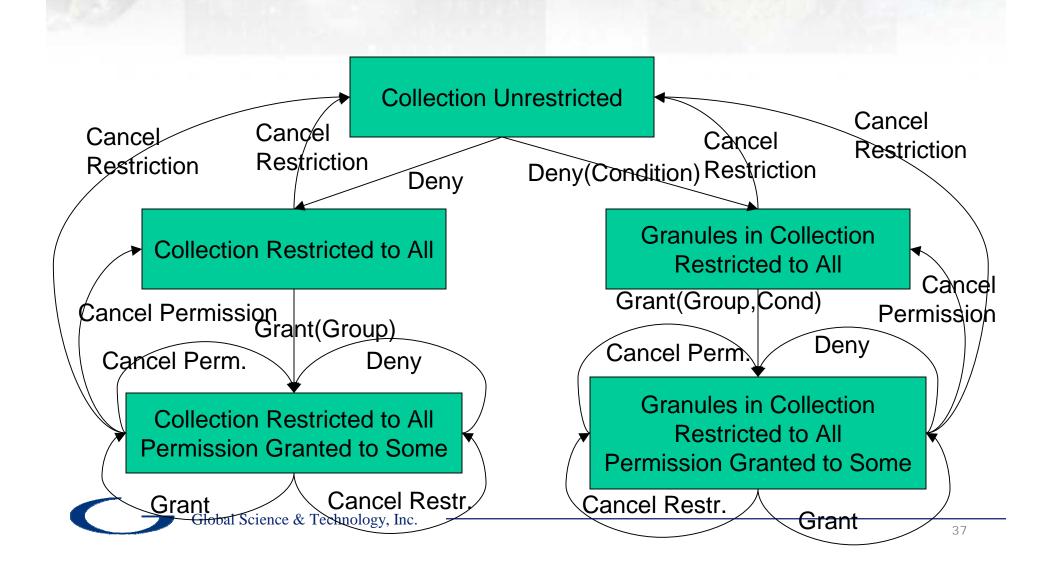


Group State Machine





DMS State Machine



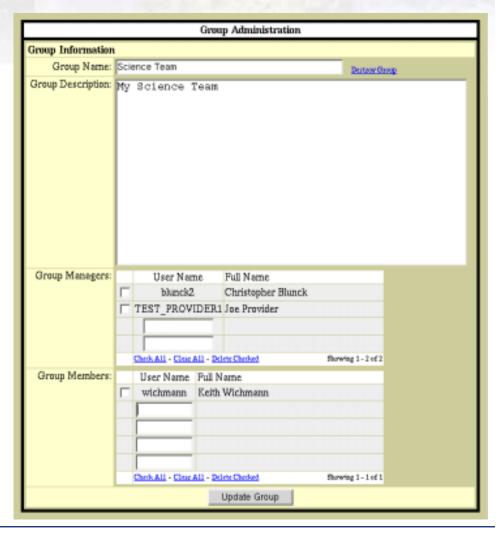
Overview of ACL GUIs

- Group Management Service
- Data Management Service
 - Condition Management
 - Data Rule Management



Group Management Service

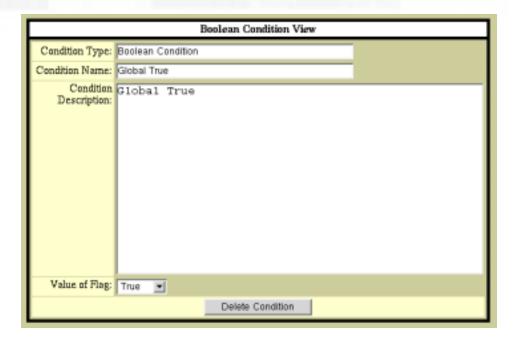
- Allows customers to manage groups
 - Group Managers
 - Group Members
- Managers can also contact group members





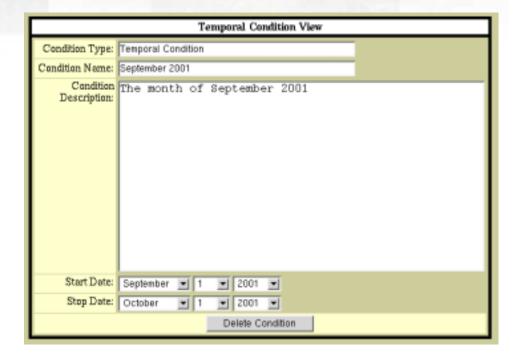
Data Management Service (Conditions)

- Condition Management is separate from Data Rule Management
- Shows how to create a globally true condition.



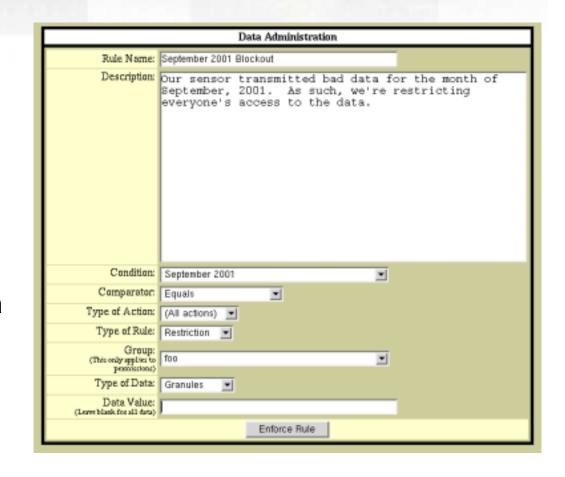
Data Management Service (Conditions)

 Shows how to create a condition based on the temporal time frame of September, 2001



Data Management Service (Restrictions)

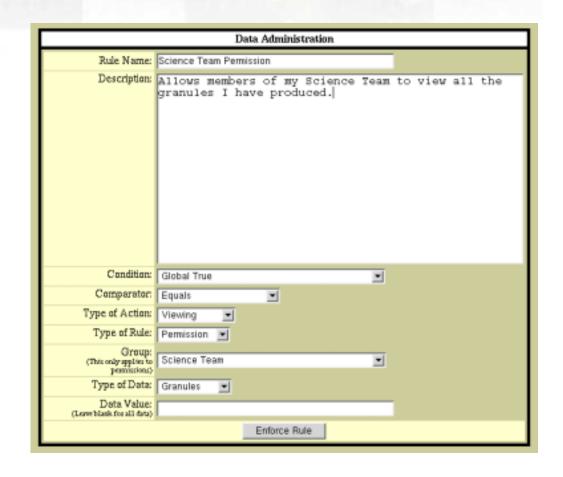
- Restrictions apply to all users.
- Restrictions are evaluated using a Condition and Comparator
- Restrictions can apply to a particular granule, or all granules (if blank)





Data Management Service (Permissions)

- Permissions apply to a particular group
- Permissions can only use Boolean conditions
- Permissions can apply to a particular granule, or all granules





User Profile Metadata

Login information:

{UserName, Password}

- No current restrictions on how the password should look.

Generic User Information:

{FirstName, LastName, EmailAddress, OptIn, OrganizationName}

- Optin denotes whether the user's information can be sent to the provider or not.

One or more addresses, each address looking like this:

{AddressID, USFormat, Street1, Street2, Street3, Street4, Street5, City, State, Zip, Country}

- AddressID is something like 'home', 'work', 'school' which uniquely describes this address.
- USFormat means the address mandates that the City, State, and Zip fields.
- AddressID, USFormat, Street1, and Country are always mandatory

One or more phone numbers, each phone number looking like this:

{PhoneName, CountryCode, AreaCode, ExchangeCode, Phone, Extension}

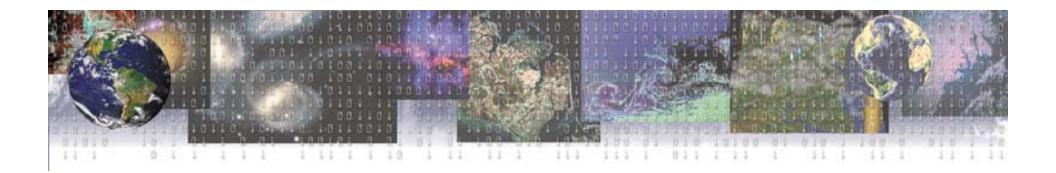
- PhoneName is similar to AddressID, and is something like 'cell', or 'home'.



Identification of ECHO User to Provider

- Currently, ECHO only uses guest access through the VO gateway
- The addition of ACLs means that ECHO Providers need to agree with ECHO as to how to identify the user
- ECHO is proposing a "Frequent Flyer" approach
 - The concept is based on travel web sites such as Expedia
 - A user can register their frequent flyer number with Expedia, which then relates it to the provider of the ticket when it brokers the order
 - ECHO could allow a user to save a userid/password, data access key, RSA identity, etc. and then use that information when brokering the order





Metadata Services

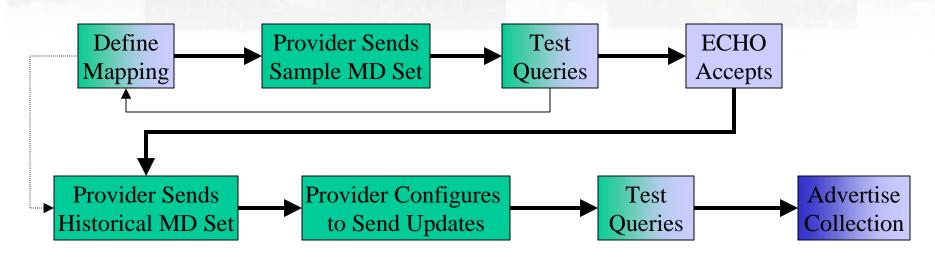


Data Model Review

- May 21 Material Review and Kickoff. (Notes)
- June 3 Science Data Model Discussion (Collections)
- (CHANGED) June 11 Science Data Model Discussion Continued (Granules, etc.)
- (CHANGED) June 18 Business (Persistent Data) Model Discussion
- (CHANGED) June 19 Application of the Data Model Provider View (Ingest)
- July 9 Application of the Data Model "Client" View (Query and Results)
- July 16-17 Application of the Business Model all views AND final commentary review and wrap-up.

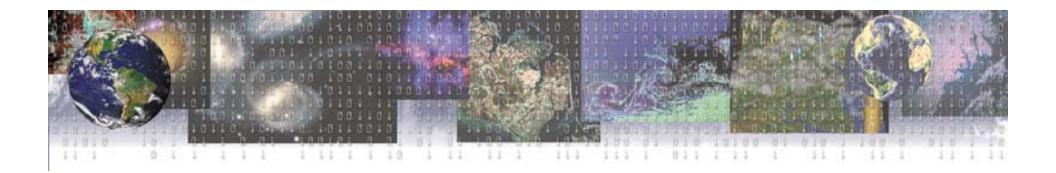


New Collection Process



Assumption: Many of a provider's mappings will be similar and the process of mapping will get easier and require less work.

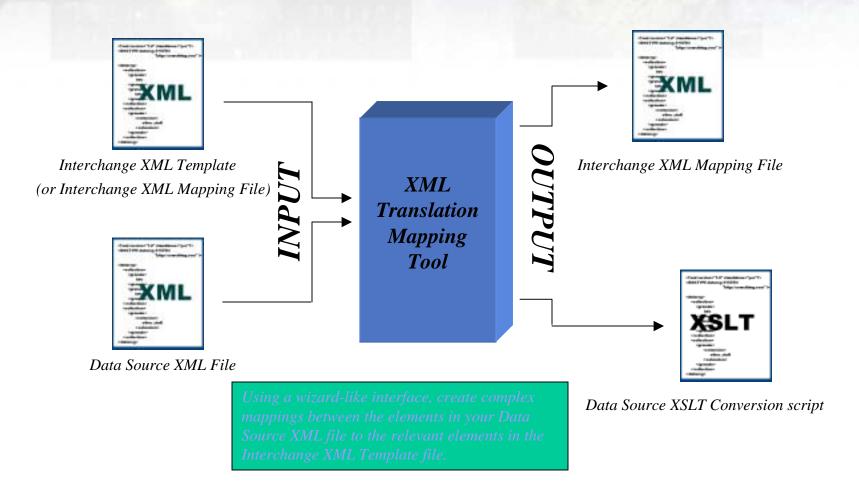




Process for Creating ECHO Compatible XML File

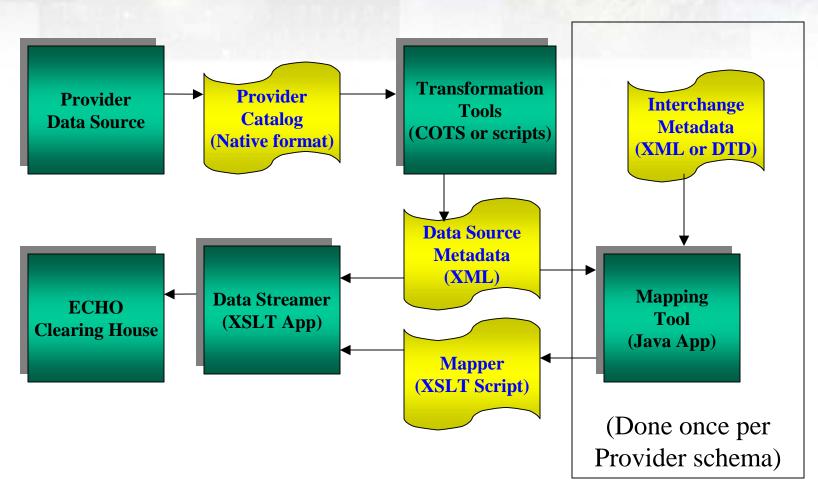


Translation Process

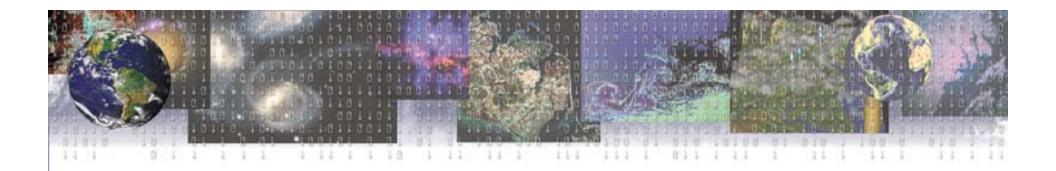




Metadata Mapping/Ingest Process







Spatial Coordinates & Projections



Spatial Coordinates and Projections

- Coordinate system Provider has choice of:
 - Flat Cartesian system
 - Geodetic Coordinate System
 - •GEOGCS ["Longitude / Latitude (WGS 84)", DATUM ["WGS 84", SPHEROID ["WGS 84", 6378137.000000, 298.257224]], PRIMEM ["Greenwich", 0.000000], UNIT ["Decimal Degree", 0.01745329251994330]]



How ECHO applies Oracle Spatial

Data types supported

- Point
- Points connected with Straight line without closure
- Gpolygon, Polygon
 - connected with straight line (great circle in geodetic coordinate system)
 - A Gpolygon or Polygon can not cover more then half of the earth for geodetic coordinate system
 - Provide enough pointers (density) to reflect the correct coverage area especially for geodetic coordinate system.
- Multi Polygon
 - no overlap between the polygons.
 - Each individual Polygon should satisfy the requirements listed above
- Circle
 - Accept latitude/longitude for center point with radius. Those parameters will be used to generate pointers to form a polygon based on geodetic coordinate system.
 - The unit of the radius must be in meter



How ECHO applies Oracle Spatial

Data types supported

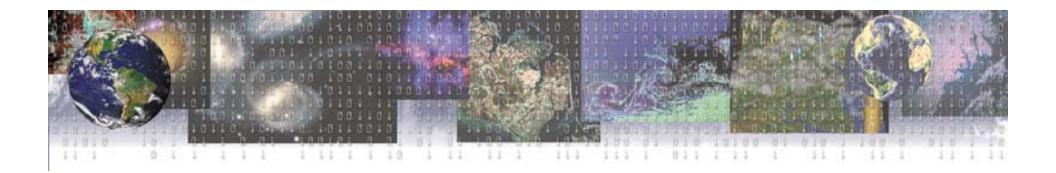
- Bounding Box
 - Based on flat coordinate system
 - No size limitation
 - No density requirement
 - •Can be converted to one or more geodetic (WGS 84) coordinate system based polygon(s) with appropriate density and stored as polygons based on geodetic (WGS 84) coordinate system.
 - However, if converted from a bounding box based on Cartesian coordinate system to polygon(s) based on geodetic coordinate system, it could cause misleading FALSE for the spatial search with certain options such as WITHIN.
- Data type conversions will be supported in the future



What is Expected from Data Providers

- Select appropriate data type for your spatial data
- For GPolygon type, put vertices in a clockwise order
- For Polygon type, put vertices in a clockwise order for both inner ring and outer ring
- The measurement unit for radius of the circle data must be in meters
- Split the GPolygon if it crosses the international day line or poles if Cartesian Coordinate system is used





Order Options



Order Line Item Options

 Currently, ECHO uses a somewhat standard set of order options for most orderable provider metadata. These order options consist mostly of packaging options like:

Media_Type, Compression, FTP_Host

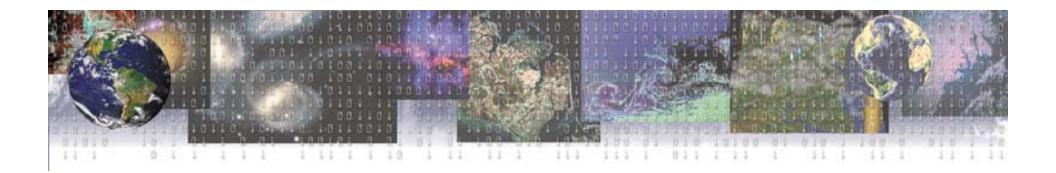
- These options are associated with each granule or collection on an individual basis.
 - To set the value of an associated option, use the 'UpdateOrderLineItem' transaction in the 'OrderEntryService'.
 - Currently, there are only options at this order-line item level. Eventually, ECHO may have options at both the provider-order, and order level.



Ingestion of Options

- Currently, any new set of options are put into the database by hand.
- However, in the future (by 5.0) we plan to do the following:
- Automate ingestion of data through BMGT file.
 - each set of options a provider supports are tagged with a unique ID
 - each granule/collection will be tagged with its option ID
- Service API for providers to manage options.
- Have a suggested 'valids' template for order options.
 - Make it easier on clients to parse the options
 - Make it easier for providers to fill out the options





Browse Services



Two Approaches to Browse

Host "science" browse in ECHO

- Pro: Quick access to HDF-EOS browse files
- Con: Hardware costs increase significantly

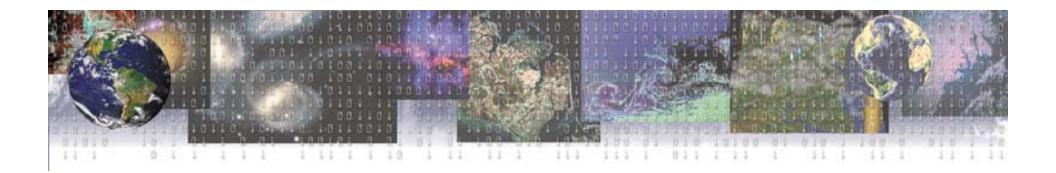
Host URL for "science" browse in ECHO

- Pro: Less ECHO Hardware
- Con: If provider is down, URL is not available

Both cases (and clients) benefit from:

- Host "browser viewable" browse in ECHO
- This would be a format viewable by standard browsers (JPG, GIF, etc.)
- Provider would define conversion
- Note that a graphic could include text used as a legend and any scales for interpreting the rest of the image
- Issue: Should there be metadata about the browse? (description, geospatial?, etc.)





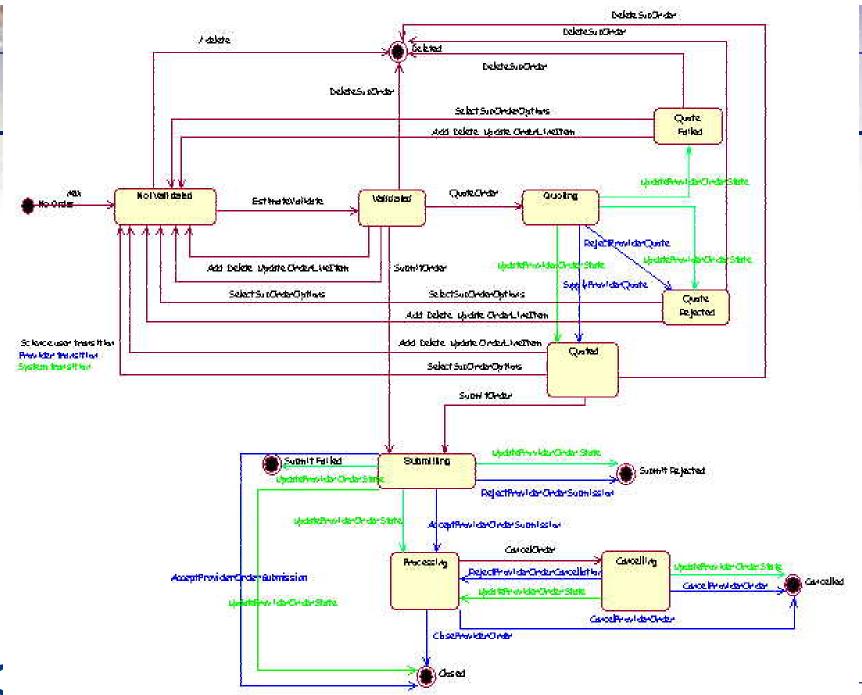
Order Interactions



Order Interactions with the Provider

- Two communication mechanisms with providers:
 - ODL over a socket connection ECS providers
 - ECHO-formatted XML over SOAP ORNL and future providers
- Three transactions create direct communications w/ the provider:
 - Submit (mandatory), Quote (SOAP only), Cancel (SOAP only)
- Two ways providers can respond back to an ECHO transaction:
 - Respond directly to the request
 - only way to respond for ODL communications
 - only 'providerTrackingID' response is mandatory for SOAP
 - Through API: ProviderOrderManagementService





ProviderOrderManagementService API

AcceptProviderOrderSubmission

CancelProviderOrder

ChangeTrackingID

CloseProviderOrder

PresentClosedOrder

Present Closed Order Summary

PresentOpenOrder

PresentOpenOrderSummary

RejectProviderOrderCancellation

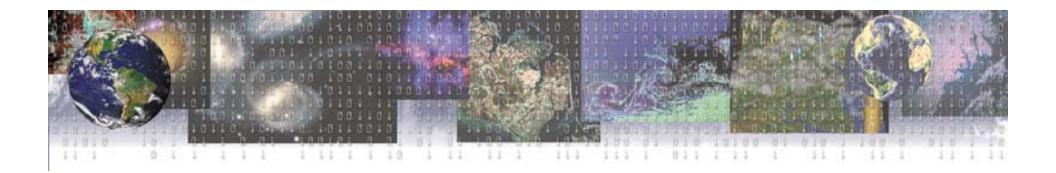
RejectProviderOrderSubmission

RejectProviderQuote

SupplyProviderQuote

UpdateStatusMessage





Data Services



Data Services

- ECHO will be adding the ability for a service provider to register a new service into ECHO
- ECHO Clients will then be able to find that service and execute it
- These services will be distributed in nature

Topics for ECHO Ops Plan

- Who identifies when a metadata update file was missed?
 - Data ingest reconciliation compare inventories at a granule level
 - Comparison is provider's responsbility, ECHO generates report of provider's holdings
- Who is involved in auditing the contents of ECHO against the contents of a provider?
- What is the operational burden of providing metadata updates to ECHO?
- Who handles problems with metadata?
 - Data reconciliation issue
 - Who is responsible to investigate, resolve and correct problems?
 - This is a longer term issue, after ingest
- Who handles requests for order status ("Where is my order?")
- Clear responsibility of ownership of the system Who shuts it off? Who takes care of it when it is misbehaving?
 - Performance monitoring and resolution
 - Ability to isolate IP addresses that are creating unworkable loads on the system
- Need a 7X24 ops plan Is there a need for weekend or night support?



- The RMA plan needs to be documented explicitly If there is a 99% goal, what are the tactics to get there?
- Responsibility for performance monitoring is there a level of service target? Will it be monitored and managed? Is it best effort? Will the addition of new providers and clients be managed in order to ensure quality of service?
- Who is responsible for the different aspects of ingest?
 - If an ingest file fails to appear at ECHO, who will notice?
 - If an ingest file that appeared at ECHO, who will notice if it is not in the clearinghouse?
- Who decides to accept a provider?
 - Science Operations Office
- Who manages the baseline of ECHO?
 - This is referencing the performance baseline
 - What about a quota for providers?
- What is the usefulness of ECHO to the provider?
- What happens when a data provider changes their data model?
- What happens when ECHO data model changes?
- We need a template ops agreement between ECHO and a provider
- · We need a template ops agreement between ECHO and a client



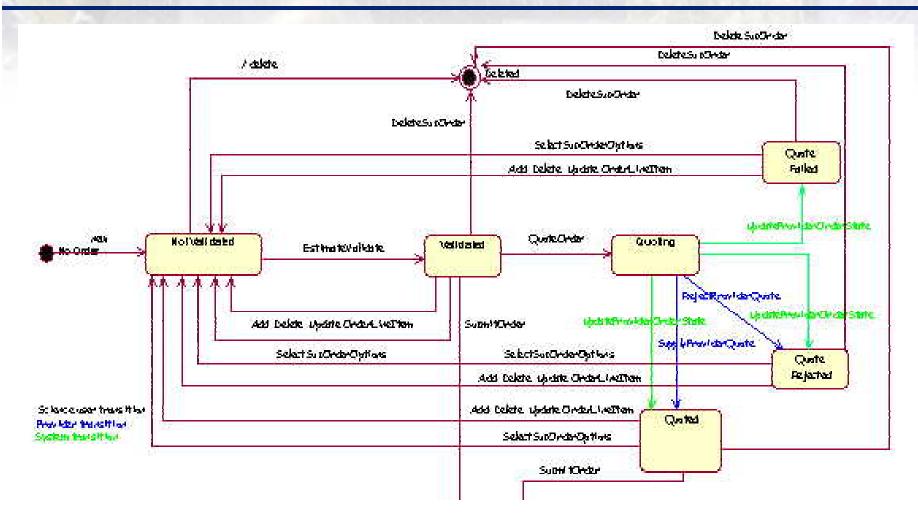
Acceptable valids

- What are the core set of attributes that are required to be put into ECHO?
- Create a metadata plan including responsibilities
- Will we use GCMD, CIP or some other source (or a mixture) for valids for various keywords in the data model?
- Some keywords may remain un-controlled
- Is there currently required information that is not really being used?

- User Services for ECHO
 - Who investigates orders that have run afoul?
- Policy on user services for clients
 - Is it a criteria for sanctioning a client?
- Is it required for a provider to have a user services presence to be accepted as an ECHO provider?
- Who do customers talk to and under what circumstances?
- Who do user services talk to and under what circumstances?
- Does ECHO allow all clients? Does ECHO provide a mechanism for blocking clients (remember there is no ironclad way to prevent...)? Does ECHO block all clients and then add them in as they are sanctioned?



Pre-submittal Order States





Post-submittal Order States

